The Annals of Statistics
1983, Vol. 11, No. 1, 129–140

ASYMPTOTIC EFFICIENCY OF THREE-STAGE HYPOTHESIS TESTS[1]

BY GARY LORDEN

California Institute of Technology

Multi-stage hypothesis tests are studied as competitors of sequential tests. A class of three-stage tests for the one-dimensional exponential family is shown to be asymptotically efficient, whereas two-stage tests are not. Moreover, in order to be asymptotically optimal, three-stage tests must mimic the behavior of sequential tests. Similar results are obtained for the problem of testing two simple hypotheses.

1. Introduction. Multi-stage hypothesis tests have obvious practical advantages over fully sequential tests in many situations and can be expected to have part of the efficiency advantage of sequential tests. Can they be asymptotically efficient when compared with sequential tests? If so, how many stages are required to do it?

The present investigation shows that in the contexts of Sequential Probability Ratio Tests (SPRT's) and Schwarz's (1962) tests for the one-dimensional exponential family, the answer is yes and that three stages are needed except in degenerate cases. The asymptotically efficient three-stage tests must necessarily imitate the behavior of asymptotically optimal sequential tests (see Remark 1 of Section 6). They can be constructed as follows. Take m observations in the first stage, and then stop if the corresponding sequential test would stop; otherwise determine a time $N_2 \geq m + 1$ to end the second stage, N_2 being designed to slightly overestimate the stopping time of the sequential test. If at N_2 one is still within the continuation region of the sequential test, then continue until time $\bar{n}$, when the third and final stage terminates. Thus, in any case the total number of observations is at least m but does not exceed $\bar{n}$. The ideas developed in the proof of the main theorem suggest that the same sort of recipe works in quite general contexts for sequential testing, such as that of Kiefer and Sacks (1963).

Multi-stage tests in the SPRT context are discussed in Section 2, and the main result for exponential families is derived in Section 3. The results of Section 4 show that two-stage tests generally have asymptotic efficiency less than one, and that asymptotically optimal tests of three or more stages have asymptotically negligible final stage. Section 5 derives some lower bounds on expected sample sizes that are refinements of an inequality of Hoeffding (1960) and are useful in the proof of Theorem 1.

2. Multi-stage competitors of the SPRT. A natural starting place for the investigation of the asymptotic efficiency of multi-stage tests is the problem of deciding which of two given densities, f and $g (f \neq g)$, is true based on independent and identically distributed observations. Let f_n and g_n denote the likelihoods after n observations and denote the usual information numbers (assumed finite) by

$$I(f, g) = E_f \log(f_1/g_1) \quad \text{and} \quad I(g, f) = E_g \log(g_1/f_1).$$

It is well-known that the optimal tests are SPRT's with expected sample sizes

$$E_f N \sim I(f, g)^{-1} \log \beta^{-1} \quad \text{and} \quad E_g N \sim I(g, f)^{-1} \log \alpha^{-1} \tag{1}$$

Received May 1981; revised September 1982.

[1] Research supported by the National Science Foundation under Grant MCS-7804870.

AMS 1970 *subject classification.* Primary 62F05; secondary 62L10.

Key words and phrases. Multi-stage hypothesis test, asymptotic efficiency.

as the error probabilities α and β go to zero.

It is easy to describe a family of two-stage tests attaining the asymptotic optimality property, (1), in the case where

$$I(f, g)^{-1}\log \beta^{-1} \sim I(g, f)^{-1}\log \alpha^{-1}. \tag{2}$$

Using the weak law of large numbers and a standard asymptotic technique (Chung, 1968, page 188), it can be shown that there exist integers $n = n(t) \geq t$ such that as $t \to \infty$

$$n = t + o(t),$$

$$P_f\{\log(f_n/g_n) < tI(f, g)\} \to 0, \quad \text{and} \quad P_g\{\log(g_n/f_n) < tI(g, f)\} \to 0. \tag{3}$$

For given α and β, choose the first stage sample size $m = n(t)$, where t is the larger of the two sides of (2). Then m is asymptotic to both sides of (2), and by (3)

$$P_f\{\log(f_m/g_m) \geq \log \beta^{-1}\} \to 1 \tag{4}$$

and

$$P_g\{\log(g_m/f_m) \geq \log \alpha^{-1}\} \to 1. \tag{5}$$

The test terminates with the first stage if the event in (4) or in (5) occurs, making the appropriate terminal decision. If neither holds, the test continues until a total of $\bar{n}$ observations have been taken and makes the same terminal decision as a fixed sample size test on $\bar{n}$ observations that is chosen to have error probabilities at most α and β. This can be done with $\bar{n} \leq Mt \leq Mm$ observations. Thus, the expected sample size under f is at most

$$m[1 + MP_f\{\log(f_m/g_m) < \log \beta^{-1}\}] = m\{1 + o(1)\} \sim I(f, g)^{-1}\log \beta^{-1},$$

and a similar result is obtained under g. Now, the probability of error under f is at most the sum of the probability that the event in (5) occurs, which is less than α, and the error probability of the test based on $\bar{n}$ observations. Hence, the error probability under f is less than 2α and, similarly, under g is less than 2β. One can achieve given α^* and β^* by going through the construction with $\alpha = \alpha^*/2$ and $\beta = \beta^*/2$, leaving the asymptotic results for expected sample sizes unaffected.

In the case where (2) does not hold but $(\log \beta^{-1})/(\log \alpha^{-1})$ is bounded away from zero and infinity, Corollary 1 of Section 4 shows that there do not exist asymptotically optimal two-stage tests. However, one can attain asymptotic optimality by using a three-stage test whose first two stages end at $\min(m_1, m_2)$ and $\max(m_1, m_2)$, where

$$m_1 \sim t_1 = I(f, g)^{-1}\log \beta^{-1}, \quad m_2 \sim t_2 = I(g, f)^{-1}\log \alpha^{-1},$$

and (4) and (5) hold for m_1 and m_2, respectively. A third stage ending at $\bar{n} \leq M \max(m_1, m_2)$ is used as before, and a similar argument shows that the expected sample sizes under f and g are both minimized asymptotically.

If the log-likelihood ratio $\log(g_1/f_1)$ has finite third moment under f and g then more explicit determinations of m_1 and m_2, of the form

$$m_i = t_i + C_i(t_i \log t_i)^{1/2}, \qquad i = 1, 2,$$

can be made along the lines of (Chung, 1968, page 214) using the Berry-Esseen theorem. In this case one obtains

$$E_fN = I(f, g)^{-1}\log \beta^{-1} + O([(\log \beta^{-1})\log \log \beta^{-1}]^{1/2}),$$

and a similar result for E_gN. The results in the next section are of this type.

It is shown in Section 4 that two-stage tests can attain the asymptotic optimality property (1) only if (2) is satisfied.

3. Three-stage tests for the exponential family. Suppose that independent and identically distributed observations $X_1, X_2, \cdots$ have density

$$f_\theta(x) = \exp\{\theta x - b(\theta)\}, \qquad \theta \in \Theta,$$

with respect to a non-degenerate σ-finite measure and Θ is the natural parameter space of the family of densities. Let $S_n = X_1 + \cdots + X_n$, $n = 1, 2, \cdots$, and note that $E_\theta S_n = nE_\theta X_1 = nb'(\theta)$ and $\mathrm{Var}_\theta S_n = n\,\mathrm{Var}_\theta X_1 = nb''(\theta)$. Sequential likelihood ratio tests of two separated hypotheses $\theta \le \theta_0$ and $\theta \ge \theta_1$, can be defined in terms of the log-likelihood functions

$$L_n(\theta) = \theta S_n - nb(\theta)$$

as follows: for given $0 < \gamma_0, \gamma_1 < 1$, stop the first time

$$\sup_{\theta>\theta_0} L_n(\theta) \ge L_n(\theta_0) + \log \gamma_0^{-1} \tag{6}$$

or

$$\sup_{\theta<\theta_1} L_n(\theta) \ge L_n(\theta_1) + \log \gamma_1^{-1}. \tag{7}$$

Relations (6) and (7) define the upper and lower stopping boundaries, respectively, of the tests arising in the study of Bayes asymptotic shapes in Schwarz (1962), and $\theta \le \theta_0$ (resp. $\theta \ge \theta_1$) is rejected only if the upper (resp. lower) boundary is crossed.

Let $\underline{\theta} < \bar{\theta}$ denote given interior points of Θ. It is shown in Lorden (1972) that if $(\log \gamma_1^{-1})/\log \gamma_0^{-1}$ is bounded away from 0 and ∞ as γ_0 and γ_1 go to zero, then the expected sample sizes of the tests defined by (6) and (7) are optimal to within $O(\log\log \gamma_i^{-1})$, $i = 0, 1$, as $\gamma_0, \gamma_1 \to 0$, uniformly for $\theta \in [\underline{\theta}, \bar{\theta}]$, among all tests with the same or smaller error probabilities. A similar result with $O([(\log \gamma_i^{-1})\log\log \gamma_i^{-1}]^{1/2})$ is proved below for a family of three-stage tests constructed to imitate the sequential tests.

To obtain the three-stage tests, it is necessary first to develop an alternative description of the stopping boundaries along the lines of Schwarz's work. If x belongs to the range of the X_i's and $x > b'(\theta_0)$, then the ray $S_n = nx$ intersects the upper stopping boundary at a point whose n-coordinate is

$$n_0(x) = \frac{\log \gamma_0^{-1}}{I_0(x)}$$

(not necessarily an integer), where

$$I_0(x) = \sup_{\theta>\theta_0}[(\theta - \theta_0)x - \{b(\theta) - b(\theta_0)\}].$$

(Define $I_0(x) = 0$ and $n_0(x) = \infty$ if $x \le b'(\theta_0)$.) For the lower stopping boundary, $I_1(x)$ and $n_1(x)$ are defined similarly. Note that $I_0 \nearrow$ in x and $I_1 \searrow$, so that $n_0 \searrow$ and $n_1 \nearrow$. Let

$$n(x) = \min(n_0(x), n_1(x)),$$

the first intersection of the ray $S_n = nx$ with one of the stopping boundaries. The function $n(x)$ characterizes the stopping region in the sense that (k, S_k) reaches or exceeds one of the boundaries if and only if $k \ge n(S_k/k)$.

Since b' is an increasing continuous function on the interior of Θ, its range is an interval (x_*, x^*). If x belongs to (x_*, x^*), then the maximum likelihood estimate of θ on $S_n = nx$ is $\hat{\theta}(x) = (b')^{-1}(x)$, and it is easily verified that the information numbers satisfy

$$I(\hat{\theta}(x), \theta_0) = I_0(x) \quad \text{if} \quad b'(\theta_0) \le x < x^* \tag{8}$$

and

$$I(\hat{\theta}(x), \theta_1) = I_1(x) \quad \text{if} \quad x_* < x \le b'(\theta_1). \tag{9}$$

It is also easy to see that there is a θ in (θ_0, θ_1) for which the information numbers satisfy

$$\frac{I(\theta, \theta_0)}{I(\theta, \theta_1)} = \frac{\log \gamma_0^{-1}}{\log \gamma_1^{-1}}, \tag{10}$$

since the left-hand side increases continuously from 0 to ∞ on (θ_0, θ_1). Let θ_2 denote the solution of (10) and let $\tilde{n} = n(b'(\theta_2))$. By (8)–(10), $n_0(x)$ and $n_1(x)$ intersect at $x = b'(\theta_2)$

where they both have the value $\tilde{n}$, so that

$$\tilde{n} = \frac{\log \gamma_0^{-1}}{I(\theta_2, \theta_0)} = \frac{\log \gamma_1^{-1}}{I(\theta_2, \theta_1)} = \max_x n(x), \tag{11}$$

using the monotonicity properties of $n_0(x)$ and $n_1(x)$. It is assumed in the main theorem that the ratio of $\log \gamma_1^{-1}$ to $\log \gamma_0^{-1}$ is bounded away from 0 and ∞, so that θ_2 remains bounded away from the endpoints of (θ_0, θ_1).

A convenient fact for use in the definition of the three-stage tests and the proof of their asymptotic optimality is the following (which is proved following the proof of the main theorem).

LEMMA 1. *There exist positive constants A and B such that if*

$$\rho_n = 1 + A(n^{-1}\log n)^{1/2},$$

then

$$P_\theta\left\{\rho_n^{-1} < \frac{n(S_k/k)}{n(b'(\theta))} < \rho_n,\ k \geq n\right\} > 1 - \frac{B}{n} \tag{12}$$

for $n \geq 1$ and $\underline{\theta} \leq \theta \leq \bar{\theta}$.

The three-stage tests and, in particular, the stopping times N_1, N_2, N_3 for the three stages can now be defined. Use $[y]$ to denote the greatest integer $\leq y$ and $\{y\}$ for the least integer $\geq y$.

Stage(i). Choose $0 < C \leq \min(I(\theta_2, \theta_0)/I(\bar{\theta}, \theta_0), I(\theta_2, \theta_1)/I(\underline{\theta}, \theta_1))$. Let $N_1 \equiv m = [C\tilde{n}]$. The restriction on C ensures that

$$m \leq n(b'(\theta)), \qquad \underline{\theta} \leq \theta \leq \bar{\theta}. \tag{13}$$

Stage(ii). Let $\bar{n} = \{\tilde{n}\}$ and set $N_2 = \min(\bar{n}, \{\rho_m^2 n(S_m/m)\})$.

Stage(iii). $N_3 \equiv \bar{n}$.

At the end of stages one and two, the test is terminated if and only if one of the boundaries is crossed, which is certainly the case after the third stage, by virtue of (11). In any case, an appropriate terminal decision is made. Note that if at the end of the first stage the test continues, then $m < n(S_m/m)$ and, hence, $N_2 > m$.

It is clear that the error probabilities, α_0 and α_1, when $\theta \leq \theta_0$ and $\theta \geq \theta_1$, respectively, satisfy

$$\alpha_0 \leq \bar{n}\gamma_0 \quad \text{and} \quad \alpha_1 \leq \bar{n}\gamma_1, \tag{14}$$

since the probability under $\theta \leq \theta_0$, for example, that (6) holds for a fixed $n \in \{1, \cdots, \bar{n}\}$ is at most γ_0, as was shown in Lorden (1972).

Let $t(a_0, a_1, \theta)$ denote the smallest possible θ-expectation of the sample size of a (possibly sequential) test with error probabilities less than or equal to a_0 and a_1.

THEOREM 1. *If $N = N(\gamma_0, \gamma_1)$ is the stopping time of the three-stage test defined above, then as γ_0 and γ_1 tend to zero*

$$E_\theta N \leq t(\bar{n}\gamma_0, \bar{n}\gamma_1, \theta) + O(((\log \gamma_i^{-1})\log \log \gamma_i^{-1})^{1/2}), \qquad i = 0, 1,$$

uniformly for $\underline{\theta} \leq \theta \leq \bar{\theta}$, provided that $(\log \gamma_1^{-1})/\log \gamma_0^{-1}$ *is bounded away from* 0 *and* ∞, *in which case $t(\bar{n}\gamma_0, \bar{n}\gamma_1, \theta)$ is of order* $\log \gamma_i^{-1}$ *for* $i = 0, 1$.

PROOF. Fix θ in $[\underline{\theta}, \bar{\theta}]$ and assume that γ_0 and γ_1 are small enough so that $m \geq 1$.

Let V denote the event in (12) when $n = m$, the first stage sample size, i.e. the event

$$\rho_m^{-1} < \frac{n(S_k/k)}{n(b'(\theta))} < \rho_m, \qquad k \geq m. \tag{15}$$

To see that

$$N \leq \{\rho_m^2 n(S_m/m)\} \quad \text{on } V, \tag{16}$$

note first that this relation holds if $N = m$, since by (13)

$$m \leq n(b'(\theta)) < \rho_m n(S_m/m) \quad \text{on } V.$$

If $N > m$, then $N_2 > m$ and either $N_2 = \bar{n}$, in which case $N = N_2$ and (16) holds by the definition of N_2, or else

$$N_2 = \{\rho_m^2 n(S_m/m)\} \geq \rho_m n(b'(\theta)) \geq n(S_{N_2}/N_2),$$

which suffices for stopping, so that $N = N_2$ and equality holds in (16). Now, (15) and (16) yield

$$N \leq \{\rho_m^3 n(b'(\theta))\} \quad \text{on } V$$

and, using Lemma 1 and the fact that $\bar{n}/m \to C^{-1}$,

$$E_\theta N \leq \{\rho_m^3 n(b'(\theta))\} + \frac{B}{m}\bar{n} = n(b'(\theta)) + O((\bar{n}\log\bar{n})^{1/2}).$$

By a modification of a lower bound of Hoeffding (1960), it can be shown (see the remark following Lemma 2 of Section 5) that

$$t(\bar{n}\gamma_0, \bar{n}\gamma_1, \theta) \geq n(b'(\theta)) - O(\bar{n}^{1/2}) \tag{17}$$

uniformly for $\underline{\theta} \leq \theta \leq \bar{\theta}$. Therefore,

$$E_\theta N - t(\bar{n}\gamma_0, \bar{n}\gamma_1, \theta) \leq O((\bar{n}\log\bar{n})^{1/2})$$

and the conclusion of the theorem follows immediately.

PROOF OF LEMMA 1. Since B may be chosen arbitrarily large, it evidently suffices to prove (12) for large n, in which case (12) is made stronger by replacing ρ_n by ρ_k. This stronger version follows from

$$P_\theta\left(\rho_k^{-1} < \frac{n(S_k/k)}{n(b'(\theta))} < \rho_k\right) \geq 1 - \frac{B'}{k^2} \tag{18}$$

by summing the complementary probabilities over $k \geq n$.

To prove (18), let J denote a closed subinterval of (x_*, x^*) containing $b'(\underline{\theta})$ and $b'(\bar{\theta})$ as interior points. Then there is a $D > 0$ such that

$$\left|\frac{n(x_1)}{n(x_2)} - 1\right| \leq D\,|x_1 - x_2| \quad \text{if} \quad x_1, x_2 \in J \tag{19}$$

by the following argument. If $x_1, x_2 \geq b'(\theta_2)$, then the left-hand side of (19) equals

$$\left|\frac{I_0(x_2)}{I_0(x_1)} - 1\right| \leq \frac{|I_0(x_2) - I_0(x_1)|}{I_0(b'(\theta_2))} \leq \frac{(\text{length of } (b')^{-1}(J))|x_2 - x_1|}{I_0(b'(\theta_2))},$$

this last as a straightforward consequence of the fact that the supremum in the definition of $I_0(x)$ is attained at $\hat{\theta}(x) = (b')^{-1}(x)$ for x in J. Relation (19) holds similarly if $x_1, x_2 \leq b'(\theta_2)$, whereas if $x_1 \leq b'(\theta_2) \leq x_2$, then (19) is obtained by noting that

$$\left|\frac{n(x_1)}{n(x_2)} - 1\right| \leq \left|\frac{n(x_1)}{n(b'(\theta_2))} - 1\right| + \left|\frac{n(b'(\theta_2))}{n(x_2)} - 1\right|,$$

which reduces the argument to the previous cases.

It suffices now to show that there exists a $Q > 0$ such that

$$P_\theta\left(\left|\frac{S_k}{k} - b'(\theta)\right| \le (Qk^{-1}\log k)^{1/2}\right) \ge 1 - \frac{2}{k^2}, \qquad \underline{\theta} \le \theta \le \bar{\theta}, \tag{20}$$

since the event in (20) implies for sufficiently large k (independent of θ) that S_k/k belongs to J, whence (19) can be applied to yield (18) for large k. To prove (20), choose $\varepsilon > 0$ such that $\underline{\theta} - \varepsilon$ and $\bar{\theta} + \varepsilon$ belong to the interior of Θ. Since b'' is continuous, there is a $Q > 0$ such that

$$| b(\theta + t) - b(\theta) - tb'(\theta)| \le \frac{Q}{8} t^2 \quad \text{if} \quad \underline{\theta} \le \theta \le \bar{\theta} \quad \text{and} \quad | t | \le \varepsilon.$$

Therefore, letting $a_k = (Qk \log k)^{1/2}$ and $t = 4(Q^{-1}k^{-1}\log k)^{1/2}$ and using Chebyshev's inequality, for sufficiently large k

$$\begin{aligned} P_\theta(S_k - kb'(\theta) \ge a_k) &\le P_\theta(\exp(tS_k - tkb'(\theta) - ta_k) \ge 1) \\ &\le (\exp(b(\theta + t) - b(\theta)))^k \exp(- tkb'(\theta) - ta_k) \\ &\le \exp\left(\frac{kQ}{8} t^2 - ta_k\right) = 1/k^2. \end{aligned}$$

A similar argument works for estimating the probability that $S_k - kb'(\theta)$ is less than or equal to $-a_k$, and putting the two results together yields (20), completing the proof of the lemma.

4. Necessity of three stages. The main theorem of this section can be used to show that three stages are, generally speaking, necessary for asymptotic optimality of tests. Results of this nature for the testing problems of the preceding sections are contained in Corollaries 1 and 2. They are based on the assertion of Theorem 3 that, barring a special relationship between the overall error probabilities, the last stage of an asymptotically optimal multi-stage test contributes a vanishingly small fraction of the total sample size.

For $k = 1, 2, \cdots$, let $C_k(\alpha, \beta)$ denote the class of k-stage tests of f vs. g having error probabilities $\le \alpha, \beta$. Define

$$t_k(\alpha, \beta, p) = \inf_{C_k(\alpha,\beta)} EN,$$

where E denotes expectation under the density p. In case $k = 1$ the tests have fixed sample size and, hence, the notation $t_1(\alpha, \beta)$ will be used to reflect the lack of dependence on p. In accord with Section 3, the notation $t(\alpha, \beta, p)$ without subscript will denote the infinimum of EN over all tests, including fully sequential ones.

For asymptotic theory, the basic fact is that

$$t(\alpha, \beta, p) \sim \min\left(\frac{\log \alpha^{-1}}{I(p, f)}, \frac{\log \beta^{-1}}{I(p, g)}\right)$$

as $\alpha, \beta \to 0$. The case where the two ratios on the right-hand side are asymptotically equal is special in that fixed sample size tests turn out to be asymptotically optimal, at least if p belongs to the exponential family generated by f and g. The hypothesis (21) of Theorems 2 and 3 is designed to rule out this special case. The next theorem shows that under this assumption at least two stages are required to minimize the expected sample size asymptotically when p is true.

THEOREM 2. *Suppose that f, g and p are densities with respect to a sigma-finite measure, that f is distinct from g and p, and that all of the information numbers are finite. If α and β satisfy*

$$\frac{\log \beta^{-1}}{\log \alpha^{-1}} \ge Q > \frac{I(p, g)}{I(p, f)} \tag{21}$$

for some $Q > 0$, *then as* $\alpha \to 0$

$$t_k(\alpha, \beta, p) \sim \frac{\log \alpha^{-1}}{I(p, f)} \quad \textit{for} \quad k \geq 2, \tag{22}$$

whereas there exist $\eta > 0$ *and* $D < \infty$ *such that*

$$t_1(\alpha, \beta) \geq (1 + \eta) \frac{\log \alpha^{-1}}{I(p, f)} - D \tag{23}$$

for all $0 < \alpha, \beta \leq 1$ *satisfying* (21).

PROOF. It suffices for (22) to show that

$$t(\alpha, \beta, p) \geq \frac{\log \alpha^{-1}}{I(p, f)} (1 - o(1)) \tag{24}$$

and that the right-hand side is attainable by two-stage tests. The latter fact is established by a construction like the one in Section 2. Choose the first-stage sample size

$$m = \frac{\log \alpha^{-1}}{I(p, f)} (1 + o(1))$$

in such a way that under p

$$P\left(\log \frac{p_m}{f_m} \geq \log \alpha^{-1}\right) \to 1$$

and, if this highly probable event fails to occur, continue until $\bar{n}$, chosen so that a fixed sample size test on $\bar{n}$ observations has error probabilities $\leq \alpha, \beta$. To prove (24), fix $0 < \varepsilon < 1$ and note that by the law of large numbers

$$\log \frac{p_N}{f_N} - (1 + \varepsilon) N I(p, f) \leq \max_n \left(\log \frac{p_n}{f_n} - (1 + \varepsilon) n I(p, f)\right) < \varepsilon \log \alpha^{-1},$$

this last with probability approaching one. Using Chebyshev's inequality,

$$P\left(\log \frac{p_N}{f_N} \leq (1 - \varepsilon) \log \alpha^{-1} \text{ and } f \text{ is rejected}\right) = P\left(\frac{f_N}{p_N} \geq \alpha^{1-\varepsilon} \text{ and } f \text{ is rejected}\right)$$

$$\leq \alpha^{\varepsilon - 1} E \frac{f_N}{p_N} 1\{f \text{ is rejected}\} = \alpha^{\varepsilon} \to 0.$$

Combining the last two relations,

$$P((1 + \varepsilon) N I(p, f) \leq (1 - 2\varepsilon) \log \alpha^{-1} \text{ and } f \text{ is rejected}) \to 0.$$

A similar result holds with f and α replaced by g and β (even if $p = g$) and, using (21), leads to

$$P((1 + \varepsilon) N I(p, f) \leq (1 - 2\varepsilon) \log \alpha^{-1} \text{ and } g \text{ is rejected}) \to 0.$$

Since either f or g must be rejected, (24) follows and (22) is proved.

The lower bound (23) on $t_1(\alpha, \beta)$ is derived by considering the exponential family of densities of the form $h = C(\theta) f^{\theta} p^{1-\theta}$, $0 \leq \theta \leq 1$.

As $\theta \to 0$, $I(h, f) \to I(p, f)$ and $I(h, g) \to I(p, g)$ by dominated convergence. By virtue of (21), then, θ can be chosen small enough so that $I(h, g)/I(h, f) < Q$.
Since (21) holds with p replaced by h, so does (24), and thus

$$t_1(\alpha, \beta, h) \geq t(\alpha, \beta, h) \geq \frac{\log \alpha^{-1}}{I(h, f)} (1 - o(1)).$$

This suffices for (23), since $I(h, f) < I(p, f)$, which holds because h lies between p and f in the exponential family.

THEOREM 3. *Under the assumptions of Theorem 2, suppose that a family of k-stage tests* ($k \geq 2$) *is given, with error probabilities* α, β *satisfying* (21). *Let* $N = N(\alpha, \beta)$ *denote the stopping time of a test in the family and* $M = M(\alpha, \beta)$ *the total sample size of its first* $k - 1$ *stages. If the tests are asymptotically optimal, i.e. if the expected sample size under* p *satisfies*

$$EN \sim \frac{\log \alpha^{-1}}{I(p, f)} \tag{25}$$

as $\alpha \to 0$, *then* $M \sim \log \alpha^{-1}/I(p, f)$ in probability (p), and $EM \sim \log \alpha^{-1}/I(p, f)$.

The proof is postponed until after the statements and proofs of the two corollaries.

COROLLARY 1. *In the problem of testing f vs. g in Section* 2, *suppose that the information numbers are finite and that* ($\log \beta^{-1}$)/($\log \alpha^{-1}$) *is bounded away from zero and infinity. Then there exist two-stage tests asymptotically minimizing both* $E_f N$ *and* $E_g N$ *if and only if* (2) *holds.*

PROOF. The construction of two-stage tests based on (3) in Section 2 suffices for the existence assertion. The converse part is proved by applying Theorem 3 with $p = g$ to conclude that the first stage sample size m must be asymptotic to ($\log \alpha^{-1}$)/$I(g, f)$, and with $p = f$ (reversing the roles of f and g in the theorem) to find m asymptotic also to ($\log \beta^{-1}$)/$I(f, g)$, whence (2) holds.

COROLLARY 2. *In the testing problem of Section* 3, *assume that* ($\log \alpha_1^{-1}$)/($\log \alpha_0^{-1}$) *is bounded away from zero and infinity. Then there do not exist two-stage tests that minimize* $E_\theta N$ *asymptotically for four distinct values of* θ *in* $[\theta_0, \theta_1]$.

REMARK. Two-stage tests *can* be asymptotically optimal for three θ-values—see Remark 3 of Section 6. But tests asymptotically optimal for an interval of θ-values must have structure similar to the tests proposed in Section 3—see Remark 1 of Section 6.

PROOF OF COROLLARY 2. Consider a sequence of (α_0, α_1)'s tending to zero and note that there is a subsequence along which ($\log \alpha_1^{-1}$)/($\log \alpha_0^{-1}$) converges to a limit $L > 0$. If $\phi_1, \cdots, \phi_4$ are distinct points in $[\theta_0, \theta_1]$, then the four numbers $I(\phi_i, \theta_1)/I(\phi_i, \theta_0)$, $i = 1, \cdots, 4$, are distinct since this ratio decreases strictly as ϕ_i goes from θ_0 to θ_1. Evidently there exist two of these four numbers that are distinct from L and lie on the same side of it—the left side, say (the other case being similar). Thus, there is a Q such that for these two ϕ_i's

$$\frac{\log \alpha_1^{-1}}{\log \alpha_0^{-1}} \geq Q > \frac{I(\phi_i, \theta_1)}{I(\phi_i, \theta_0)}$$

for sufficiently small α_0, α_1 in the subsequence. By Theorem 3, asymptotic optimality at the ϕ_i's requires that the first-stage sample size be asymptotic to ($\log \alpha_0^{-1}$)/$I(\phi_i, \theta_0)$ for *both* ϕ_i's, which is impossible since the two information numbers are not equal.

PROOF OF THEOREM 3. Clearly

$$N \sim \frac{\log \alpha^{-1}}{I(p, f)} \text{ in probability } (p), \tag{26}$$

since N is at least this large in probability by the argument leading to (24) and cannot be asymptotically larger with positive probability, since EN is by hypothesis asymptotic to the right-hand side of (26).

The key to the proof is to consider the conditional error probabilities, $\tilde{\alpha}$ and $\tilde{\beta}$ (say), given the first M observations, of the last stage, which always has fixed sample size (possibly zero) and is therefore subject to (23). Given $0 < \varepsilon < \frac{1}{4}$, Chebyshev's inequality yields

$$P\left(\tilde{\alpha}\frac{f_M}{p_M} > \alpha^{1-\varepsilon}\right) \le \alpha^{\varepsilon-1}E\left(\tilde{\alpha}\frac{f_M}{p_M}\right) = \alpha^{\varepsilon-1}E_f\tilde{\alpha} = \alpha^{\varepsilon},$$

which goes to zero as $\alpha \to 0$, so that

$$\log \tilde{\alpha}^{-1} + \log\frac{p_M}{f_M} \ge (1-\varepsilon)\log \alpha^{-1} \qquad \text{w.pr.} \to 1. \tag{27}$$

By the law of large numbers,

$$\begin{aligned}\log(p_M/f_M) - (1+\varepsilon)MI(p,f) &\le \max_n(\log(p_n/f_n) - (1+\varepsilon)nI(p,f)) \\ &\le \varepsilon \log \alpha^{-1} \qquad \text{w.pr.} \to 1,\end{aligned}$$

and by (26)

$$MI(p,f) \le NI(p,f) \le 2\log \alpha^{-1} \qquad \text{w.pr.} \to 1,$$

which combine to yield

$$\log(p_M/f_M) - MI(p,f) \le 3\varepsilon \log \alpha^{-1} \qquad \text{w.pr.} \to 1.$$

Using this last relation with (27) leads to

$$\log \tilde{\alpha}^{-1} \ge [(1-4\varepsilon)\log \alpha^{-1} - MI(p,f)]^+ = \log \tilde{\alpha}_1^{-1}, \tag{28}$$

say, with probability approaching one. Arguing similarly and using (21),

$$\log \tilde{\beta}^{-1} \ge [(1-4\varepsilon)\log \beta^{-1} - MI(p,g)]^+ = \log \tilde{\beta}_1^{-1},$$

say, and thus

$$N - M \ge t_1(\tilde{\alpha}, \tilde{\beta}) \ge t_1(\tilde{\alpha}_1, \tilde{\beta}_1) \qquad \text{w.pr.} \to 1. \tag{29}$$

Using (21), $\log \tilde{\beta}_1^{-1}$ is seen to be at least $Q \log \tilde{\alpha}_1^{-1}$ so that (23) applies to the extreme right-hand member of (29) and, hence,

$$N - M \ge (1+\eta)\frac{\log \tilde{\alpha}_1^{-1}}{I(p,f)} - D \qquad \text{w.pr.} \to 1.$$

Replacing N by $(1+\varepsilon)(\log \alpha^{-1})/I(p,f)$, which by (26) is an upper bound in probability, and using (28) to replace $\log \tilde{\alpha}_1^{-1}$ leads after simplification to

$$M + D\eta^{-1} \ge (1 - 4\varepsilon - 5\varepsilon\eta^{-1})\frac{\log \alpha^{-1}}{I(p,f)} \qquad \text{w.pr.} \to 1.$$

Since ε can be arbitrarily small, $(\log \alpha^{-1})/I(p,f)$ is thus asymptotically a lower bound in probability on M and, using Chebyshev's inequality, on EM also. Since $M \le N$, (25) and (26) show that the same quantity is an asymptotic upper-bound in both senses and the theorem is proved.

5. Lower bounds on expected sample sizes. The lower bound (17) used in the proof of Theorem 1 is easily established for $\theta_1 \le \theta \le \bar{\theta}$, since $t(\bar{n}\gamma_0, \bar{n}\gamma_1, \theta)$ is at least the expected sample size of an SPRT of θ vs. θ_0, which is at least $(\log(\bar{n}\gamma_0)^{-1})/I(\theta, \theta_0) - o(1)$. The case $\underline{\theta} \le \theta \le \theta_0$ is similar, and the case $\theta_0 < \theta < \theta_1$ is a straightforward consequence of (32). Inequalities (30)–(32), which may be useful in other contexts, are extensions of inequality (1.4) of Hoeffding (1960). These inequalities can be "solved for EN" to obtain inequalities more closely resembling Hoeffding's.

LEMMA 2. *Suppose that N is the stopping time of a test of f_0 against f_1 having error probabilities at most $Q\gamma_i$, $i = 0, 1$ where $Q \ge 1$. Let p be a density under which*

$$I_i = E\log(p(X)/f_i(X))$$

and

$$\sigma_i^2 = \operatorname{Var}\log(p(X)/f_i(X)), \qquad i = 0, 1,$$

are positive and finite. Then

$$\min_i \frac{\log\gamma_i^{-1}}{I_i} \le EN + \left(\min_i \frac{\log\gamma_i^{-1}}{I_i}\right)\left\{\frac{1}{2}\left(\frac{\sigma_0}{\log\gamma_0^{-1}} + \frac{\sigma_1}{\log\gamma_1^{-1}}\right)(EN)^{1/2} + \frac{\log 2Q}{\min_i\log\gamma_i^{-1}}\right\}, \tag{30}$$

$$\min_i \frac{\log\gamma_i^{-1}}{I_i} \le EN + \frac{1}{2}\left(\frac{\sigma_0}{I_0} + \frac{\sigma_1}{I_1}\right)(EN)^{1/2} + \frac{\log 2Q}{\min_i I_i}, \tag{31}$$

and

$$\min_i \frac{\log\gamma_i^{-1}}{I_i} \le EN + \frac{\rho(\gamma_0,\gamma_1)}{\max_i I_i}\left\{\frac{1}{2}(\sigma_0+\sigma_1)(EN)^{1/2} + \log(2Q)\right\}, \tag{32}$$

where $\rho(\gamma_0, \gamma_1) = (\max_i \log\gamma_i^{-1})/\min_i\log\gamma_i^{-1}$.

REMARK. In the application of (32) to derive (17), $p = f_\theta$ and the left-hand side of (32) becomes $n(b'(\theta))$. Letting $Q = \bar{n}$, EN can be replaced by $t(\bar{n}\gamma_0, \bar{n}\gamma_1, \theta)$, which is at most $\bar{n}$. The right-hand side of (32) is then at most $t(\bar{n}\gamma_0, \bar{n}\gamma_1, \theta) + O(\bar{n}^{1/2})$, since $\rho(\gamma_0, \gamma_1)$ is bounded above by hypothesis, $\max_i I(\theta, \theta_i)$ is bounded away from zero, and σ_1 and σ_2 are at most $(\bar{\theta} - \underline{\theta})$ times the maximum of $(b''(\theta))^{1/2}$ on $[\underline{\theta}, \bar{\theta}]$.

PROOF OF LEMMA 2. Let p_n denote $p(X_1)\cdots p(X_n)$ and define f_{0n} and f_{1n} similarly. Defining the events $A_i = \{p_N > 0 \text{ and } f_i \text{ is rejected}\}$ and using Wald's well-known argument,

$$Q\gamma_1 \ge P_1(A_1) = E\frac{f_{1N}}{p_N}1\{A_1\}. \tag{33}$$

Assuming without loss of generality that

$$\gamma_1 \ge \gamma_0, \tag{34}$$

let $\lambda = \log\gamma_1^{-1}/\log\gamma_0^{-1} \le 1$.

Using the obvious analog of (33),

$$Q\gamma_0^\lambda \ge (Q\gamma_0)^\lambda \ge \left(E\frac{f_{0N}}{p_N}1\{A_0\}\right)^\lambda \ge E\left(\frac{f_{0N}}{p_N}\right)^\lambda 1\{A_0\},$$

and, hence,

$$2Q\gamma_1 = Q\gamma_0^\lambda + Q\gamma_1 \ge E\left(\left(\frac{f_{0N}}{p_N}\right)^\lambda 1\{A_0\} + \frac{f_{1N}}{p_N}1\{A_1\}\right)$$

$$\ge E\min\left(\left(\frac{f_{0N}}{p_N}\right)^\lambda, \frac{f_{1N}}{p_N}\right).$$

Taking logarithms, changing signs, and using Jensen's inequality,

$$\begin{aligned}\log\gamma_1^{-1} - \log 2Q &\le E\max\left(\lambda\log\frac{p_N}{f_{0N}}, \log\frac{p_N}{f_{1N}}\right)\\ &\le \max(\lambda I_0, I_1)EN + E\max(\lambda Z_N^{(0)}, Z_N^{(1)}),\end{aligned} \tag{35}$$

where

$$Z_n^{(i)} = \log\left(\frac{p_n}{f_{in}}\right) - nI_i, \qquad n = 1, 2, \cdots.$$

Write

$$\max(\lambda Z_N^{(0)}, Z_N^{(1)}) = \tfrac{1}{2}(\lambda Z_N^{(0)} + Z_N^{(1)}) + \tfrac{1}{2}\,|\,\lambda Z_N^{(0)} - Z_N^{(1)}\,|$$

and use Wald's equations for first and second moments to get

$$\begin{aligned} E\max(\lambda Z_N^{(0)}, Z_N^{(1)}) &= \tfrac{1}{2}\, E\,|\,\lambda Z_N^{(0)} - Z_N^{(1)}\,| \le \tfrac{1}{2}[E(\lambda Z_N^{(0)} - Z_N^{(1)})^2]^{1/2} \\ &= \tfrac{1}{2}(EN)^{1/2}(\operatorname{Var}(\lambda Z_1^{(0)} - Z_1^{(1)}))^{1/2} \le \tfrac{1}{2}(EN)^{1/2}(\lambda\sigma_0 + \sigma_1). \end{aligned}$$

Using this in (35) and then dividing by $\log \gamma_1^{-1}$ yields

$$1 - \frac{\log 2Q}{\log \gamma_1^{-1}} \le \left(\max_i \frac{I_i}{\log \gamma_i^{-1}}\right) EN + \frac{1}{2}(EN)^{1/2}\left(\frac{\sigma_0}{\log \gamma_0^{-1}} + \frac{\sigma_1}{\log \gamma_1^{-1}}\right),$$

which, after dividing by the coefficient of EN and using (34), leads to (30). Relations (31) and (32) both follow easily from (30).

6. Additional remarks.

1. It is natural to ask whether the three-stage tests of Section 3 must mimic Schwarz's sequential tests in order to be asymptotically optimal. The answer according to Theorem 2 is yes, to a substantial degree they must. Assuming for convenience that $(\log \gamma_1^{-1})/(\log \gamma_0^{-1})$ converges to a positive limit, so that $\theta_2 \to \theta^*$, say, Theorem 2 requires that if the true θ is outside an arbitrarily small neighborhood of θ^* then the second stage stops at a time asymptotic to $n(b'(\theta))$ in probability. It is easily shown that the first stage sample size, m, say, must go to infinity, so that $n(b'(\hat{\theta}_m))$ is also asymptotic to $n(b'(\theta))$, and thus the second stage stopping time must be chosen in a way that is asymptotically equivalent to the method of Section 3, i.e. essentially dictated by Schwarz's stopping region. More leeway is clearly permissible in the choice of the first and third stage stopping times, although Theorem 2 does imply that the probability of a non-negligible third stage (e.g. $\ge$ a fraction ε of the total sample size) goes to zero for θ outside a neighborhood of θ^*. (It is straightforward to extend this to all θ by using the fact that any test can be truncated at $\bar{n}$ without significantly increasing its error probabilities.)

2. An associate editor pointed out how different these three-stage tests are from the asymptotically optimal three-stage confidence estimation rules of Hall (1981), which rarely finish with the second stage and seem to bear little resemblance to fully sequential procedures. As the preceding remark clarifies, the difference seems to be based on fundamental differences between the testing and estimation problems.

3. Corollary 2 of Section 4 is best possible in the sense that there do exist two-stage tests with asymptotically minimum sample size for three θ-values under special conditions. In fact, the construction of two-stage competitors of the SPRT in Section 2 suffices. Suppose that (2) holds and p is the density belonging to the exponential family generated by f and g that satisfies $I(p, f)I(f, g) = I(p, g)I(g, f)$. Then the construction of Section 2, with the second stage (when needed) stopping at a fixed $\bar{n}$ asymptotic to $(\log \alpha^{-1})/I(p, f)$, works and yields asymptotically minimum expected sample size at p (as well as f and g), since $\bar{n}$ is asymptotically minimum for p.

4. The referee raised a question about higher-order improvements in the asymptotic optimality stated in Theorem 1 that might be achieved by using more than three stages. By a fairly straightforward though tedious extension of the arguments in Section 3, it can

be shown that in $k \geq 3$ stages one can achieve the minimum expected sample sizes to within $t^r(\log t)^{1-r}$, where t is the sequential minimum and r is ½ to the power "number of stages minus two". Thus, each additional stage can reduce the extra sampling by roughly a square root factor, but it is doubtful that this kind of improvement is worth the extra complexity of tests with more than three stages.

Acknowledgment. Thanks are due to a referee for careful reading and criticism and to Bob Berk, Peter Bickel, and David Siegmund for helpful discussions.

REFERENCES

CHUNG, K. L. (1968). *A Course in Probability Theory.* Harcourt, Brace, and World, New York.
HALL, PETER (1981). Asymptotic theory of triple sampling for sequential estimation of a mean. *Ann. Statist.* **9** 1229–1238.
HOEFFDING, W. (1960). Lower bounds for the expected sample size and the average risk of a sequential procedure. *Ann. Math. Statist.* **31** 352–368.
KIEFER, J. and SACKS, J. (1963). Asymptotically optimum sequential inference and design. *Ann. Math. Statist.* **34** 705–750.
LORDEN, G. (1972). Likelihood ratio tests for sequential k-decision problems. *Ann. Math. Statist.* **43** 1412–1427.
SCHWARZ, G. (1962). Asymptotic shapes of Bayes sequential testing regions. *Ann. Math. Statist.* **33** 224–236.

DEPARTMENT OF MATHEMATICS
CALIFORNIA INSTITUTE OF TECHNOLOGY
SLOAN LABORATORY OF MATHEMATICS AND PHYSICS
PASADENA, CALIFORNIA 91125